Analyzing Different Interaction Modes for Crowdsourcing

Aastha Goyal
School of Information
University of Texas at Austin
aastha.goyal@utexas.edu

Rupali Roy
School of Information
University of Texas at Austin
rupali.roy@utexas.edu

December 12, 2019

Abstract

We aim to compare different interaction modes that can be used in crowdsourcing to analyze them on the basis of effectiveness, accessibility and reliability. In this project, we harness the power of interactive conversational interfaces provided by smart speakers to develop a voice based interface and compare it to the traditional web-based interface for crowdsourcing. First we designed a web based task for recognizing the happiness level of a person shown in the image. We used COCO data set to get the images and its corresponding caption. We also developed a voice-based crowdsourcing task that was deployed on Amazon Alexa, which can be considered as an accessible voice based marketplace for visually impaired people. We conducted a qualitative analysis on these two interaction modes and compared them on the basis of accessibility, efficiency and reliability. Through our qualitative analysis of the two interaction modes we found that using smart speaker based crowdsourcing tasks have several advantages and hence we provide recommendation for designing crowdsourcing marketplaces for visually impaired people thereby reducing the population bias.

1 Introduction

Motivation The overall aim of this project is to analyse and compare different interaction modes present for crowdsourcing on the basis of accessibility to a larger and varied set of population. Engaging a diverse population holds value in lives of both the people who have difficulty finding jobs, and the quality of work by reducing the bias in population. We focus on people with disabilities, for whom it is convenient to access crowdsourcing platforms at their comfort but are constrained by inaccessible crowd work or difficulty finding the accessible tasks.

Previous Work Prior work on accessibility has noted that crowdsourcing could offer a potential source of employment for physically impaired people but they lose upon employment opportunities due to inaccessible crowd work [7]. Further, there have been some studies proposing involvement of voice-based methods in crowdsourcing task to make it accessible to a blind audience [6]. Hettiachchi et al. suggested the potential of smart speakers in creating accessible crowd work [4]. We extend this idea to compare the efficacy of these interaction methods to make crowdsourcing more inclusive.

Idea We aim to harness the advancements in technology to create accessible crowdsourcing tasks and compare it to traditional tasks. In order to create accessible content we will design two different tasks using two different interaction modes - a web-based task hosted on Amazon Mechanical Turk and a voice-based tasks deployed on Amazon Alexa. We will then compare the user responses with each other and standard tasks to find which methodology is best suited to make crowdsourcing more accessible.

2 Related Work

2.1 Reduce Population Biasing

Diversifying the crowd worker population is one of the important ways to make crowdsourcing platform more inclusive. An online survey was conducted on AMT, to develop a relationship between per worker hourly wage and worker demographics like country of residence, gender and health conditions. This survey suggested that people with disabilities contribute significantly to crowdsourcing tasks [2]. Crowdsourcing platforms can be a good source of income for physically challenged people as they can complete the task at their convenience. There is some previous work trying to involve people with ASD (Autism Spectrum Disorder) in image transcription tasks [3] and visually impaired people in voice transcription tasks [6]. Our work is different because it focuses on comparing the different methodologies to find the best suited mode for making crowdsourcing more inclusive and reducing the population biasing.

2.2 Creative crowdsourcing interaction modes

Majority of the current crowdsourcing tasks are based on the traditional ways of uploading web based tasks on platforms like Amazon Mechanical Turk. There are many crowdsourcing tasks like transcription [6] that involve audio or speech data for which we can use voice-based interfaces. Another study proposes to ensure the compliance with WCAG guidelines so that people using features like talk back can easily understand it [7]. Further, a study by Het-tiachchi et al. proposed the potential of smart speakers in creating accessible crowd work [4]. We are extending this idea to harness the power of robust interactive speech recognition in smart speakers for crowdsourcing and comparing it to other interaction modes.

2.3 Improving Web Accessibility

When creating any web based tasks, it is important to ensure the UI elements are labeled properly, HTML headings are properly placed and the landmarks such as ARIA are used wherever required [1]. The lack of these essential UI elements in any website prohibits the screen reading software to work properly, hence making it difficult for people with disability to perform the tasks with good accuracy. The WCAG 2.0 guidelines is used to verify the accessibility of websites, it ensures that the website has required HTML labels and tags that are used by screen reader. Prior surveys suggest that very few crowdsourcing task follow these guidelines [7] [5]. In our work we create crowdsourcing tasks that adhere to the WCAG 2.0 guidelines and compare it to voice-based methodologies.

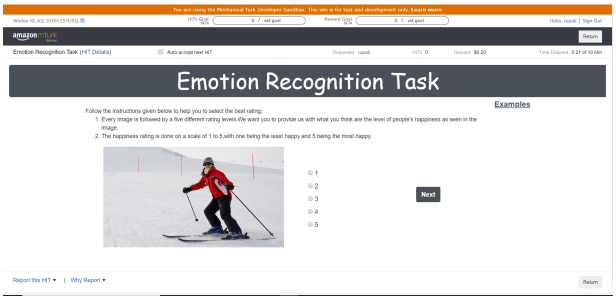

Figure 1: Design of the AMT Crowdsourcing Task

3 Methods

3.1 Designing AMT crowdsourcing task

We have designed a web-based crowdsourcing task hosted on Amazon Mechanical Turk to collect a the happiness rating of an image on the scale of 5. The task consists of series of randomly selected images from the COCO data set that contain at least one person. We have only used the images with humans as it is easier to map emotions to humans. The crowd worker is asked to interpret the happiness of the person seen in the image and rate it on a scale of 1 to 5, 5 being the happiest and 1 being the least happy. The interface design of the AMT task is as shown in figure 1 and 2.

3.2 Creating Smart Speaker Skill for voice based crowdsourcing task

With increasing popularity, smart speakers like Google Home and Amazon Alexa are widely available in every

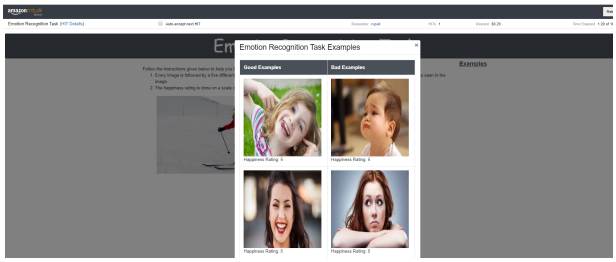

Figure 2: Example Images for AMT Crowdsourcing Task

Figure 3: Architecture of Alexa Application

household. They have made conversational interfaces easily available that can be exploited for crowdsourcing. We aim to make use of this advancement in technology to our benefit by making crowdsourcing more accessible and available to a diverse audience. By means of our application, people can conveniently perform crowdsourcing tasks along with their regular household work.

The architecture of our Alexa application is shown in the figure 3. A user can initialize the crowdsourcing task by saying the phrase "Alexa open Survey Me". On initialization, Alexa gives detailed instructions on how to perform the task followed by a series of questions that need to be answered by the crowd worker. The task consists of list of image caption taken from the COCO data set. The crowd worker needs to listen to the image captions carefully and then tell the happiness rating on the scale of 5. Then, Alexa converts users responses to text and sends users response to our service. Our service then stores the user response and their id in S3 database. Alexa follows up with the next question and waits for user response.

4 Experimental Design

In our first experiment we asked various crowd workers to work on the two different interaction modes. We used the images and corresponding captions from COCO data-set to design the experiment. The images were used in the web-based task and the captions for the same image were used in the voice-based task hosted on Alexa. The crowd worker was required to perform both the crowdsourcing tasks and we used the results to measure the efficiency of voice-based interface.

Efficiency: We measure the efficiency of our voice-based model by comparing its user responses from web-based interface. In our experiment we calculated the average value of happiness rating given by user and the variance in their responses.

In the second and third experiment, our main purpose is to perform a qualitative analysis between the traditional web based and the voice-based crowdsourcing task on the basis of user experience. After completing the crowdsourcing task on the two interaction modes. We conducted a survey using google forms amongst the users to understand their personal experiences with respect to the following metrics:

1. Accessibility

   The accessibility is defined as how easily and readily available any particular interaction mode is for a varied set of population. For the web based crowdsourcing task, we checked the accessibility with respect to the smooth execution of task on different types of browser. We asked the users through the survey, if they faced any issues while executing the task on different browsers or they faced any issues loading the web page. For the voice-based task, we consider a broader aspect of accessibility, we measure accessibility in terms of the availability of the smart speaker based task to larger section of people.

2. Reliability

   In our experiment we define reliability in terms of the ability to accurately record the happiness rating for the web based and smart speaker based task. In case of voice-based task, it may be possible that multiple users try to interact with a single smart speaker and

A woman posing for the camera standing on skis.

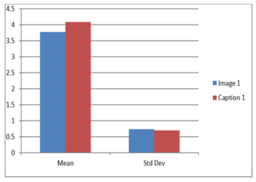

Figure 4: Experimental Result for image 1

A male tennis player in white shorts is playing tennis

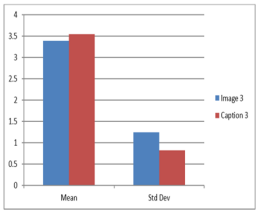

Figure 6: Experimental Result for image 2

A couple of baseball players standing on the field.

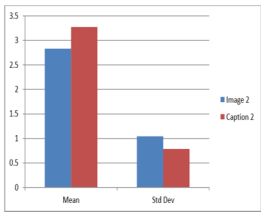

Figure 5: Experimental Result for image 2

it may happen that it fails to accurately record the happiness rating.

5 Experimental Results

5.1 Efficiency

Functionality Using smart speaker for crowdsourcing task gives several advantages like greater accessibility, improved convenience and reduction of population bias but there are some major drawbacks as well. One of the important drawback is that the smart speaker can mainly be used for task that can be completed using voice recognition. In case of tasks that include image segmentation, smart speakers cannot be used as they do not have any touch interface.

5.2 Accessibility

Voice-based interfaces like smart speaker allow more users to leverage the benefits of crowdsourcing. People with disabilities like vision or physical impairment can also participate in crowdsourcing tasks thereby receiving its economic benefits and reducing the bias in crowd worker population.
Smart speakers allow crowd workers to work at their convenience without having to interact with any web or mobile interface. Crowd workers can interact with voice-based task on smart speakers while working on other regular things. While it is an advantage, it is accompanied by the condition of availability. Web portals are readily available at anywhere and anytime but this might not be true with smart speakers for everyone.

5.3 Reliability

We compare the variance and standard deviation of the happiness rating of both these modes and can conclude that voice based task have comparable results to that of the web based tasks as seen in figure 4, 5 6. Though smart speakers employ natural conversational interfaces, the automatic speech recognition techniques for identifying the input speech is prone to errors. Various scenarios like multiple user interaction or different accents have been the reason for error in recorded user response. This error is unlikely to occur in traditional web based crowdsourcing task that use the keyboard input for getting the answers. Thus, we conclude that web-based crowdsourcing task are more accurate and reliable as compared to smart speakers.

5.4 Future Work

Visual Interfaces In our current system, the only crowd sourcing tasks that can be implemented for smart speakers are voice-based or speech related tasks due to the limitation on visual interfaces. With advancement in technology, smart speakers are evolving. Amazon has introduced visual interface for smart speakers, Echo show, which can be used in further development of crowdsourcing tasks on smart speakers. We have created a basic task for smart speakers with no touch interface but this can be extended to a variety of crowd sourcing task developed on smart interfaces.

6 Conclusion

Although the crowdsourcing platforms like Amazon Mechanical Turk have provided more than half a million people across the globe with additional employment opportunities, there are still several accessibility barriers for people with visual impairment to access the traditional web based crowdsourcing platform . After analyzing and comparing the results from both the interaction modes, we found that on the basis of accessibility; the voice based crowdsourcing task is accessible to a varied set of population and thus it helps us in reducing population biasing. In terms of reliability, when we compared the output of the experiments that were conducted in these two interaction modes, we found that web based task is more reliable and accurate as there are more chances of error in the smart speaker based task when multiple users interact with the system. Efficiency of web based task is comparatively more as it has the capability to host image segmentation task which is not possible in the Amazon Alexa that was used in our project as it lacks visual display.

References

[1] R. Calvo, S. K. Kane, and A. Hurst. Evaluating the accessibility of crowdsourcing tasks on amazon's mechanical turk. In *Proceedings of the 16th international ACM SIGACCESS conference on Computers & accessibility*, pages 257–258. ACM, 2014. 2

[2] K. Hara, A. Adams, K. Milland, S. Savage, B. V. Hanrahan, J. P. Bigham, and C. Callison-Burch. Worker demographics and earnings on amazon mechanical turk: An exploratory analysis. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, page LBW1217. ACM, 2019. 2

[3] K. Hara and J. P. Bigham. Introducing people with asd to crowd work. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 42–51. ACM, 2017. 2

[4] D. Hettiachchi, N. van Berkel, T. Dingler, F. Allison, V. Kostakos, and J. Goncalves. Enabling creative crowd work through smart speakers. 1, 2

[5] S. Swaminathan, K. Hara, and J. P. Bigham. The crowd work accessibility problem. In *Proceedings of the 14th Web for All Conference on The Future of Accessible Work*, page 6. ACM, 2017. 2

[6] A. Vashistha, P. Sethi, and R. Anderson. Bspeak: An accessible voice-based crowdsourcing marketplace for low-income blind people. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page 57. ACM, 2018. 1, 2

[7] K. Zyskowski, M. R. Morris, J. P. Bigham, M. L. Gray, and S. K. Kane. Accessible crowdwork?: Understanding the value in and challenge of microtask employment for people with disabilities. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pages 1682–1693. ACM, 2015. 1, 2